

Extraction and modelling of three-dimensional Urban Objects from VHR Satellite Stereo Imagery

Thomas Krauß

DLR, Remote Sensing Institute, Münchener Str. 20, 82234 Oberpfaffenhofen, Germany

ABSTRACT

The work presented in this paper shows the possibility of an automatic extraction of three dimensional urban objects from very high resolution (VHR) satellite scenes from anywhere of the world. Actual VHR satellites like GeoEye, World-View-2 or 3 or the Pliades system have ground sampling distances (GSD, “pixel sizes”) of 0.3 to 0.7 centimetres. All these systems allow also the acquisition of in-orbit-stereo-images. These are two or more images of the same location on ground acquired in the same orbit of the satellite from different viewing angles mostly only some seconds apart. From such stereo or – if more than two images were acquired – multistereo images in a first step a high resolution digital surface model (DSM) can be extracted with the same GSD as the stereo imagery. In the second step the inevitable errors and holes in the generated DSM will be filled and corrected using the multispectral imagery. Beneath the very high resolution panchromatic images which are used for the generation of the DSM also lower resolution – normally about 1/4 of the resolution of the panchromatic bands – multi-spectral images are acquired. These contain at least the four visible/NIR (VNIR) bands blue, green, red and near-infrared (NIR). Some sensors have more VNIR bands like World-View-2 (coastal, blue, green, yellow, red, red-edge and two NIR bands) or even additionally short-wave-infrared (SWIR) bands like World-View-3. From these mutispectral bands in a third step a spectral classification can be derived. This classification is used mainly for discrimination of vegetation and non-vegetation areas and the detection of water areas. The last step in this pre-processing comprises the correct orthorectification of the DSM and the pan-sharpened multispectral image. After this pre-processing of the stereo-imagery urban objects like buildings, trees, roads, bridges, and so on can be detected and in a last step these objects will be modeled to produce a final object-model of the satellite-scene or parts of it. In this paper the method is described and applied to an example satellite imagery.

Keywords: 3D Remote Sensing, Urban Objects, DSM correction, Classification

1. INTRODUCTION

Since the launch of the Ikonos- or QuickBird-satellites in 1999 or 2001 respectively the 1-m-limit of image resolution (ground sampling distance, GSD) from space was crossed and many new applications arose. Newest very high resolution (VHR) satellites deliver already imagery with GSDs of 0.5 down to 0.3 metres. But with increasing ground sampling distances also the scene sizes shrink or the amount of data grows rapidly. In this work such VHR imagery will be used for deriving high resolution digital surface models (DSM) from stereo or multi-stereo imagery of such sensors. Table 1 gives a short overview of actual VHR satellites, and their capabilities.

In this paper we present a method for automatically extracting (simple) 3D city models from only a VHR multispectral stereo scene. So we need sensors providing a high resolution panchromatic (PAN) band beneath at least four multispectral bands blue, green, red and near infrared (NIR). EROS-B and WorldView-1 mentioned in table 1 can not be used due to the lack of multispectral bands. For 3D reconstruction the acquisition of at least two satellite images in one orbit is mandatory. Better will be three images: one in nadir and the other two acquired with off-nadir angles of e.g. ± 15 degree (see fig. 1). But this latter acquisition geometry is offered only for Pléiades. Since we need at least two images from one orbit also most of the archive data can not be used due to the acquisition of mostly only single and no stereo scenes.

Further author information: (Send correspondence to Thomas Krauß)
E-mail: thomas.krauss@dlr.de

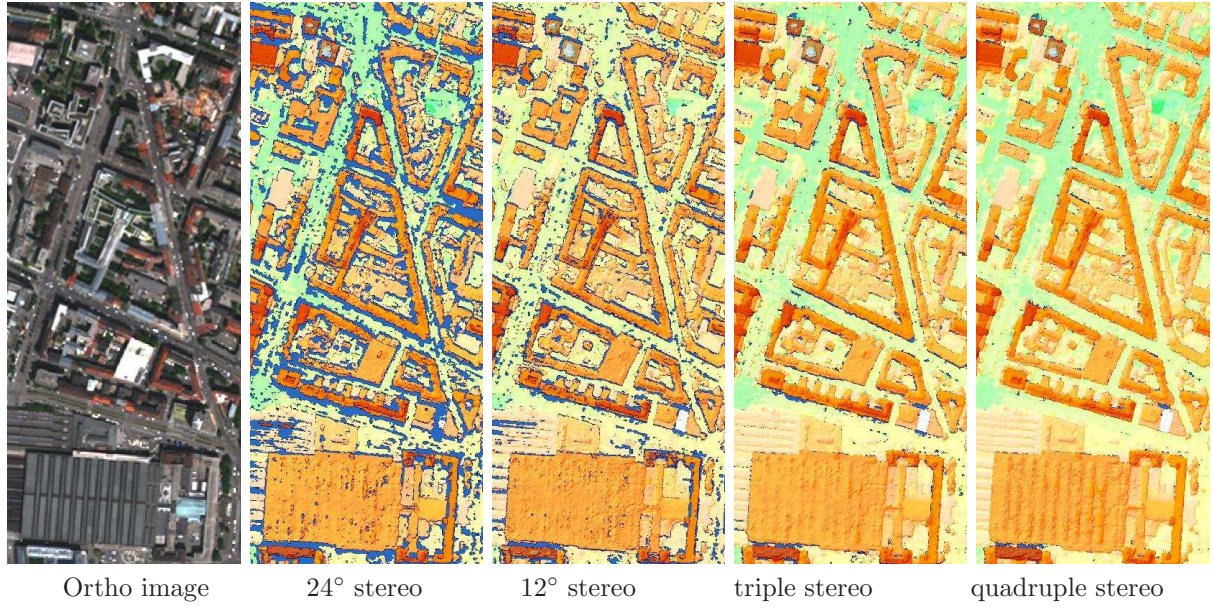


Figure 1. Influence of stereo angle and number of stereo images – example: four acquisitions of WorldView-2 over Munich
Table 1. Overview of VHR satellites usable for deriving high resolution DSMs, the prices are for multispectral 4 band stereo imagery (if available), Ikonos and QuickBird acquire no more new data, only archive-data is available, Pléiades allow as only provider also triple stereo scenes as standard acquisition mode (information compiled from: www.euspaceimaging.com, www.apollomapping.com, www.geo-airbusds.com, www.satimagingcorp.com, directory.eoportal.org at 2016-03-07)

| | Ikonos-2 | QuickBird | EROS-B | WV-1 | GeoEye-1 | WV-2 |
|-------------------------------|-----------------------|-----------------------|-----------------------|------------------------|-----------------------|-----------------------|
| Launch (–End) | 1999–2015 | 2001–2015 | 2006 | 2007 | 2008 | 2009 |
| GSD Panchromatic | 0.82 m | 0.73 m | 0.7 m | 0.5 m | 0.41 m | 0.46 m |
| GSD Multispectral | 3.28 m | 2.90 m | - | - | 1.64 m | 1.80 m |
| Multispectral bands | 4 | 4 | - | - | 4 | 8 |
| SWIR bands | - | - | - | - | - | - |
| Price/km ² Archive | 20 USD | 35 USD | 7.50 USD | 28.50 USD | 39.50 USD | 39.50 USD |
| Min Area Archive | 210 km ² | 25 km ² | 25 km ² | 100 km ² | 100 km ² | 100 km ² |
| Price/km ² New | - | - | | 92.50 USD | 105.50 USD | 105.50 USD |
| Min Area New | | | | 100 km ² | 100 km ² | 100 km ² |
| Accuracy CE90 | 15 m | 20 m | 9 m | 4 m | 4 m | 3.5 m |
| Acquisition area | 11×11 km ² | 17×17 km ² | 25×20 km ² | 51×112 km ² | 15×15 km ² | 16×16 km ² |
| Dynamic range | 11 bits | 11 bits | 10 bits | 11 bits | 11 bits | 11 bits |

WorldView-3¹ provides beside the 30 cm panchromatic band also the same eight multispectral bands like WorldView-2 does (coastal, blue, green, yellow, red, red-edge, NIR-1, NIR2) but with an GSD of 1.2 m.

For our examples and illustrations in this paper we use a WorldView-2 stereo pair over Munich (Germany) acquired on 2010-07-12 at 10:30:16 and 10:30:55 respectively. The azimuth and incidence angles of the two scenes of the stereo pair were 13.9/194.5 and 5.2/17.6 degree. So the first (aka “left”) scene was acquired from north, the second from south. The sun-azimuth and -elevation were 155 and 62.2 degree. For the example a section of 3000 m × 3000 m with a ground sampling distance of 0.5 m around the center of Munich was choosen as shown in fig. 2.

Table 2. Overview of VHR satellites usable for deriving high resolution DSMs (continued)

| | Pléiades | KompSat-3 | WV-3 |
|-------------------------------|------------------------|------------------------|-----------------------|
| Launch (–End) | 2011 | 2012 | 2014 |
| GSD Panchromatic | 0.7 m | 0.7 m | 0.31 m |
| GSD Multispectral | 2.8 m | 2.8 m | 1.24 m |
| Multispectral bands | 4 | 4 | 8 |
| SWIR bands | - | - | 8 |
| Price/km ² Archive | 52 USD | 16 USD | 46 USD |
| Min Area Archive | 25 km ² | 25 km ² | 100 km ² |
| Price/km ² New | | | 118.00 USD |
| Min Area New | | | 100 km ² |
| Accuracy CE90 | 3 m | 48.5 m | 3.5 m |
| Acquisition area | 350×20 km ² | 16×200 km ² | 13×13 km ² |
| Dynamic range | 12 bits | 14 bits | 11 bits |

2. METHOD

2.1 Overview

Nearly all methods for city modeling are based on an existing DSM and a fitting (true) ortho photo. But DSMs generated from VHR satellite imagery mostly suffer from many holes and blunders due to mismatches and occlusions. So if a DSM is calculated from VHR stereo satellite imagery the holes have to be filled. If there are errors in filling the holes in turn the ortho image which is calculated based on this DSM will be distorted in these areas.

To overcome such problems we present here a new approach. In our method of DSM generation in a first step a so called height map (HM) fitting exactly on one of the original input images is generated (without loss of generality subsequently we use the “left” or “first” stereo image). This height map contains the same holes as usual in VHR DSMs. But it can now be filled using the (multispectral) information from the original image. Only after this filling the DSM and in turn the true ortho image is generated.

The following steps in the model generation are as usual based on DSM and ortho imagery. First the digital terrain model (DTM) and the normalized DEM (nDEM) containing only the heights of objects above the ground are derived from the DSM. Afterwards a spectral classification is applied to the pansharpened ortho image. Based on the nDEM and the classification the object detection and object extraction is performed. Finally the extracted objects are modeled and output as 3D object model.

2.2 Height map generation using SGM

For deriving the height map (HM) we use the so called semi global matching (SGM) developed at DLR by [2](#) and adapted and improved for VHR stereo imagery.^{[3](#)} SGM is a dense stereo matching method which delivers for each pixel of the input stereo image pair one height value (or a no data value if the matching failed). VHR satellite images like the ones listed in [tab. 1](#) are accompanied usually with standardized so called rational polynomial coefficients (RPCs, [4, 5](#)) replacing individual physical sensor models. These coefficients define functions for samples (x-coordinate in the original image) and lines (y-coordinate in the original image) depending on longitude (λ), latitude (φ) and the ellipsoidal height h over the WGS84 ellipsoid as

$$x = \frac{f_{smp,num}(\lambda, \varphi, h)}{f_{smp,den}(\lambda, \varphi, h)} \quad \text{and} \quad y = \frac{f_{line,num}(\lambda, \varphi, h)}{f_{line,den}(\lambda, \varphi, h)}. \quad (1)$$



Figure 2. Panchromatic input image sections 6000×6000 px, $3000 \text{ m} \times 3000 \text{ m}$, left and right stereo, Munich, acquired at 2010-07-12

Each of the functions $f()$ are polynomials of third order in λ , φ and h with 20 coefficients. So over all a RPC is defined by 80 coefficients and 10 scaling and offset parameters for x , y , λ , φ and h .

In a first step characteristic points between the two input stereo images – we use the panchromatic image pair as shown in fig. 2 due to the higher resolution – are searched and a bundle adjustment is carried out to fit the images with an accuracy better than 0.5 pixels onto each other. In general a small shift of the images of about 5 m is sufficient. From the listed sensors only QuickBird imagery needs an affine image-to-image transformation in this step.

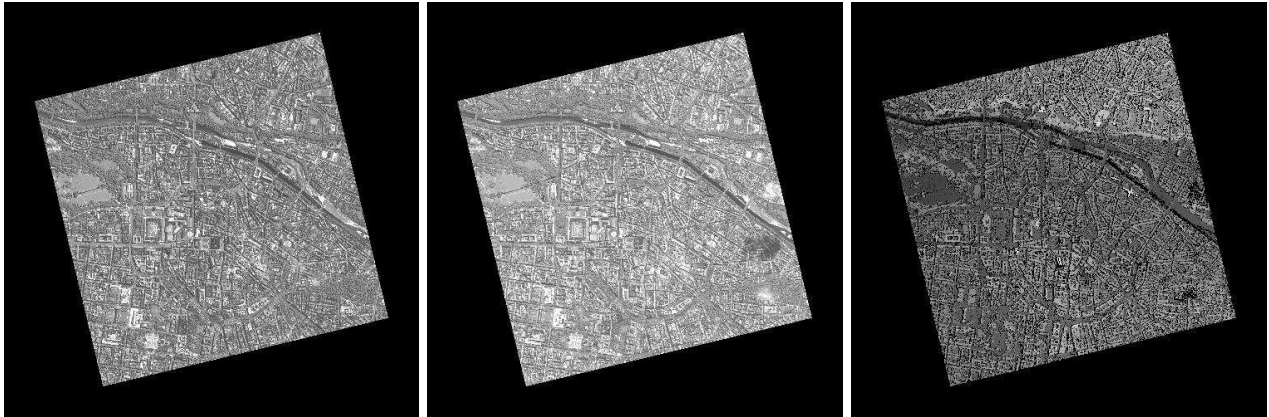


Figure 3. Transformed epipolar images (left, center) and derived disparity image (right)

The SGM method derives for each pixel in an image the distance to it's stereo mate in the other image. So the result is a so called disparity image fitting exactly on one of the two input images (without loss of generality we assume from now on always the first or "left" image as stated already above). Each pixel contains the distance to the stereo mate in the other image – the disparity d – or a no-data value if the stereo matching failed – e.g. the pixel can not be seen in the second image. To perform this pixelwise matching the two images of the stereo pair have to be in epipolar geometry as shown in fig. 3. In this special geometry the disparities due to height

differences are only horizontal shifts between the images.

Calculating geographic coordinates for a pixel in the left image is done by iterative solution of the rational polynomial function in eq. 1 for different heights h and in turn the x and y coordinates of these in the second image and vice versa give the directions of the epipolar line at this point. Repeating this for all points gives the epipolar lines for the whole image. Satellite imagery is despite of the line scanner acquisition geometry very stable so it's sufficient to do the calculation only for a sparse grid of points.⁶ For numerical and speed reasons we tile the input images and process the epipolar transformation and the semi global matching on each tile pair. The result of the SGM is the so called disparity map containing for each pixel in the left epipolar image the pixel-distance (disparity) d of this pixel to it's stereo mate in the right epipolar image as shown in fig. 3, right.



Figure 4. Height map fitting on left input image, 6000×6000 px, $3000 \text{ m} \times 3000 \text{ m}$

Back transformation of the disparity maps from epipolar geometry and applying the RPCs (eq. 1) gives the corresponding height-map tile in ellipsoidal heights as shown in fig. 4 fitting exactly on the left input image. It is calculated by forward intersection using the RPC functions for known (x_l, y_l) and (x_r, y_r) which are the transformed back epipolar coordinates (x, y) and $(x + d, y)$ and varying h providing for each pixel the geographic coordinates λ , φ and the ellipsoidal height h .

2.3 Height map correction

Since the height map calculated as described above suffers from many mismatches and occlusions a sophisticated filling has to be developed. Occlusions are areas which can not be seen in both of the stereo images and such no matching can be performed and these areas appear also as no-data in the height map. Fig. 5 (center) shows clearly these no-data-areas in a small detail of fig. 4 – the occlusions are the void areas north of buildings, the other void (black) areas are failed matches.

For filling the height map we use the information from the undistorted original image which fits exactly on the calculated height map (remember: we constructed the height map in this way). For more information content we calculate in a first step the pan sharpened multispectral image by combining the high spatial resolution of the panchromatic image with the eight multispectral TOA (top of atmosphere) bands of lower resolution following 7. As described in 8 the original, left multispectral WorldView-2 image shown in fig. 6 (left) is transformed to top of atmosphere (TOA) reflectances as shown in fig. 6 (right) using gain and offset values from the metadata

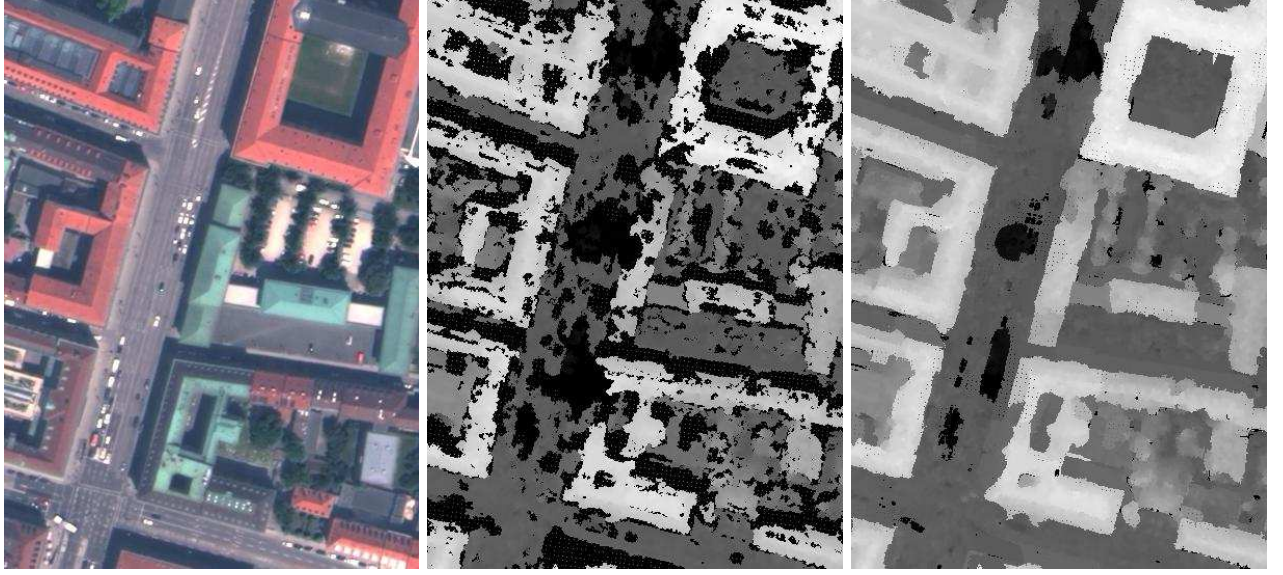


Figure 5. Details 200 m \times 275 m of pansharpened TOA image (left), unfilled height map (center) and filled height map (right)

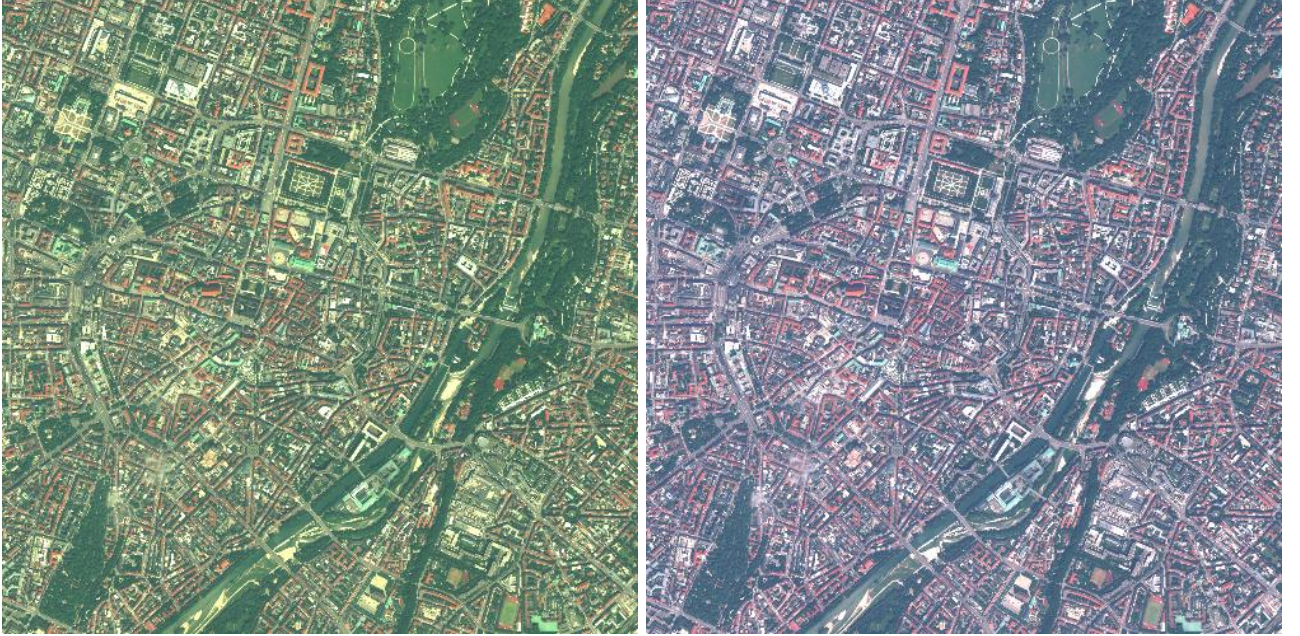


Figure 6. Multispectral image, bands 5, 3, 2 (red, green, blue) image sections 1500 \times 1500 px or 3000 m \times 3000 m, left: radiances, right: top of atmosphere reflectances

to transform the digital numbers in the image to physical radiance units and the acquisition date and the solar zenith angle for TOA calculation. The resulting TOA reflections are values between 0 and 100 %.

As described in 9 for each no-data pixel in the height map $H(x, y)$ an array $\rho(x, y)$ containing the pansharpened TOA reflectances for each of the eight bands is fetched. Afterwards in a 9×9 window $(i, j) = ([-4, 4], [-4, 4])$ around this pixel the best 5 pixels satisfying $\sqrt{\sum_{b=1}^8 (\rho_b(x, y) - \rho_b(x + i, y + j))^2} < 2.5 \%$ and containing non-no-data values in $H(x + i, y + j)$ are selected and the median of the corresponding 5 $H(x + i, y + j)$ is filled to $H(x, y)$. This procedure is iteratively repeated ten times for the whole height map. The result of the filling is shown in fig. 5 (right). The afterwards still remaining holes are filled by applying a median filter with

a small radius of 2 px iteratively until all voids are filled – non void areas keep their original values and are not affected by the filtering.

2.4 DSM generation and ortho image generation

In the next step the filled height map is transformed to a DSM by simply applying the iterative recursion of eq. 1 to calculate (λ, φ, h) from the pixel coordinates (x, y) and the height value h at (x, y) in the height map for each pixel. If more than one height value would be projected to the same DSM pixel the highest value is chosen. This happens in cases vertical walls get matched in the height map. All vertical pixels in the wall will share the same geo-coordinate and so the same pixel in the DSM but bear different heights.

Due to the orthorectification occluded areas like such on the bottom of walls which are invisible for the first satellite image keep no-data values. These areas are filled using the iterative median filtering approach as with the finally height map filling described above.

Afterwards the pansharpened original image is orthorectified using the filled DSM.

2.5 DTM generation

Based on the geographic coded DSM (see fig. 7, left) the digital terrain model (DTM, see fig. 7, center) representing the ground terrain can be calculated using a simple morphologic method. The default radius used for the morphologic filtering is 100 m. For the 0.5-m-DSM this corresponds to a radius of 200 pixels. For speed reasons first the DSM is scaled down by a factor $1/40$ using the minima of each 40×40 pixels area. Afterwards a quantile filter returning the 10-%-quantile for each 9×9 window (radius $5 \times \text{scale } 40 = 200\text{px}$) is applied followed by a 90-%-quantile filter of the same radius. These two quantile filters represent a kind of stable morphological opening (erosion followed by dilation) more insensitive toward positive and negative outliers. Finally a gaussian filter with $\sigma = 2.5\text{ px}$ (half the radius) is applied and the result scaled up back to the original size.



Figure 7. Reprojected digital surface model (DSM, left), digital terrain model (DTM, center) and normalized DEM (nDEM, right) as $\text{DSM} - \text{DTM}$

Fig. 7 (right) shows the normalized digital elevation model (nDEM) calculated by subtracting the DTM from the DSM. The nDEM represents the heights of elevated objects above the ground and range in the example case Munich from -10 to 60 m.

2.6 Spectral Classification

For object detection first a spectral classification based on the orthorectified TOA image has to be performed (TOA reflectances in $\% \cdot 10$, so a value of 10 represents 1 %). The classification follows the fuzzy-approach described in 8. For WorldView-2 and WorldView-3 imagery the bands blue (B), green (G), red (R), NIR1 (N), yellow (Y) and red-edge (E) (band 2, 3, 5, 7, 4 and 6) are used. Further the fuzzy-operators as defined in tab. 3 are used.

Table 3. Fuzzy-operators

| Symbol | Name | Definition |
|--------------------|--------------------|--|
| $a \gtrsim_d b$ | Fuzzy greater-than | $\begin{cases} 0 & \text{for } a < b - d \\ \frac{a-b+d}{2d} & \text{for } b - d \leq a \leq b + d \\ 1 & \text{for } a > b + d \end{cases}$ |
| $a \lesssim_d b$ | Fuzzy lower-than | $\begin{cases} 1 & \text{for } a < b - d \\ \frac{b-a+d}{2d} & \text{for } b - d \leq a \leq b + d \\ 0 & \text{for } a > b + d \end{cases}$ |
| $a \tilde{\cup} b$ | Fuzzy or | $\max(a, b)$ |
| $a \tilde{\cap} b$ | Fuzzy and | $\min(a, b)$ |

First so called down- and up-indicators d and u ranging from 0 to 1 are calculated:

$$\begin{aligned}
d &= (B \gtrsim_5 G + 35) \tilde{\cap} (G \gtrsim_{14} Y + 16) \tilde{\cap} (Y \gtrsim_5 R + 15) \tilde{\cap} (R \gtrsim_5 E + 10) \tilde{\cap} (E \gtrsim_{2.5} N + 7.5) \tilde{\cap} (N \lesssim_{10} 110) \\
u &= ((B \lesssim_{2.5} G - 2.5) \tilde{\cap} (G \lesssim_{2.5} R - 27.5) \tilde{\cap} (G \lesssim_5 N - 100)) \tilde{\cup} \\
&\quad ((B \lesssim_5 G - 5) \tilde{\cap} (G \lesssim_5 Y - 5) \tilde{\cap} (Y \lesssim_5 R - 5) \tilde{\cap} (R \gtrsim_5 E + 5) \tilde{\cap} (E \gtrsim_5 N + 5))
\end{aligned} \tag{2}$$

The down-indicator d is a measure for falling reflectance values with increasing wavelengths $B > G > Y > R > E > N$ and $N < 110$. Vice versa the up-indicator u is a measure for raising reflectance values with increasing wavelengths $B < G < R < N$ or $B < G < Y < R > E > N$. Next the normalized digital vegetation index (NDVI) ν is calculated and the vegetation mask V and the water mask W as probability-values between 0 and 1 are derived:

$$\nu = \frac{N - R}{N + R} \quad ; \quad V = (\nu \gtrsim_{0.1} 0.2) \quad ; \quad W = (d \gtrsim_{0.05} 0.65) \tilde{\cup} ((\nu \lesssim_{0.0125} -0.1875) \tilde{\cap} (N \lesssim_{50} 250)) \tag{3}$$

The soil mask S is simply the above derived up-indicator $S = u$. Fig. 8 (left) shows the derived spectral classification for the Munich test area. The water mask is shown in blue, the vegetation mask in green and the soil mask in red. The color intensities reflect the probability values – black is 0 %, full color is 100 %.

2.7 Object detection

For the object detection the spectral classification is joined with a binary height mask \hat{H} derived from the nDEM containing all objects higher than a certain threshold – in our case 5 m. Also the fuzzy water (W), vegetation (V) and soil (S) probabilities in the spectral classification are thresholded by 50 % to give the according binary masks \hat{W} , \hat{V} and \hat{S} (\neg denoting “not”) as shown in fig. 8 (right):

- Water: \hat{W} (blue)
- Trees: \hat{V} and \hat{H} (dark green)
- Grass: \hat{V} and $\neg \hat{H}$ (light green)
- Buildings: $\neg \hat{V}$ and $\neg \hat{W}$ and \hat{H} (red)
- Road/soil: $\neg \hat{V}$ and $\neg \hat{W}$ and $\neg \hat{H}$ (gray)

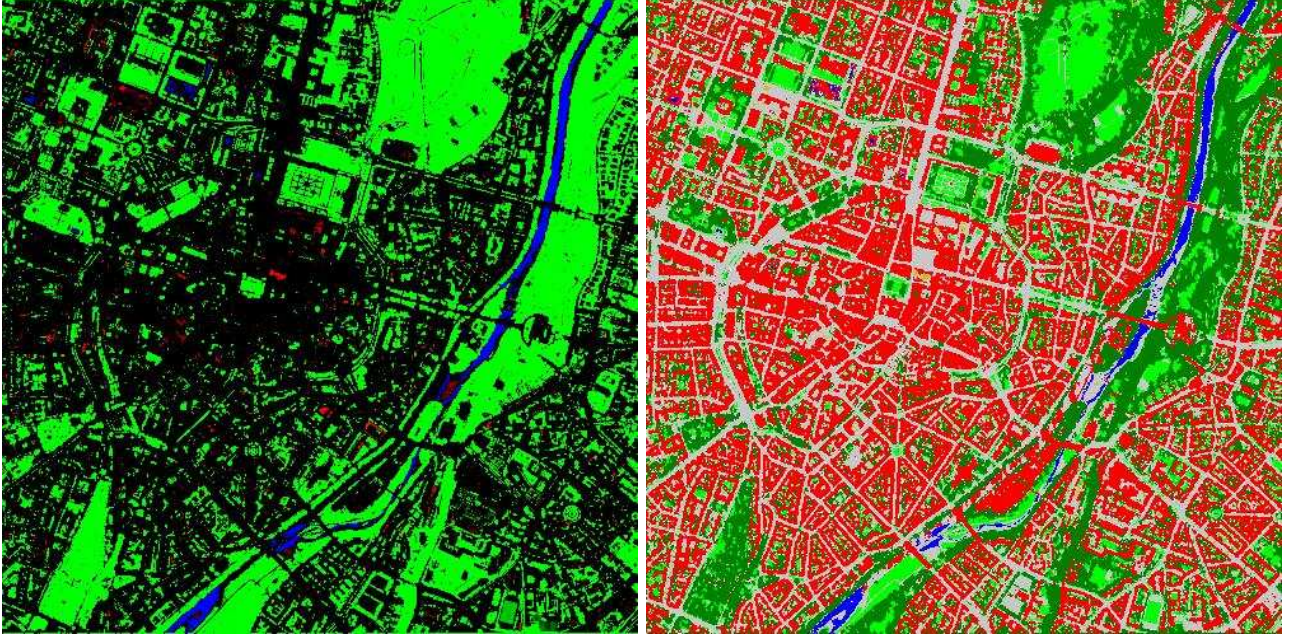


Figure 8. Spectral classification (left, blue: water, green: vegetation, red: bare soil) and detected objects (right, blue: water, light green: grass, dark green: trees, red: buildings, gray: roads/soil)

2.8 Object extraction and modeling

The modeling is based on the derived object detection and the heights from the nDEM. The resulting model depends on resolution and accuracy of input data. From VHR satellite data only levels of detail (LOD, see CityGML¹⁴) 0, 1 and 2 can be derived, which represent a simple 2.5D DSM (optionally textured) as LOD0, polygonal models with flat roofs as LOD1 and parametric object models with refined outer structure (e.g. roof shapes) as LOD2.

2.8.1 Trees

For extracting and modeling trees first the “high vegetation” – the dark green areas in fig. 8 (right) – which denote the trees are used to mask the nDEM. This masked nDEM is smoothed using a gaussian filter. Applying a inverted watershed transformation modified for speed we gain areas belonging to a local maximum in the tree-area nDEM – the crown tops of the trees. Filtering the tree-areas for size and compactness gives for the test area 41118 tree objects. For modeling the trees simply the maximum height of the filled DSM inside the modeled tree area is choosen. Fig. 10 (left) show the modeled tree objects in green overlayed on the DSM. In fig. 9 for a small section the panchromatic ortho image, the classification and the automatically extracted tree- and building-models are shown for more clarity.

2.8.2 Buildings

For extracting a vectorized building mask first all segments classified as buildings (red) in fig. 8 (right) are selected and after appropriate morphological filtering for each of the segments the outline coordinates are used to estimate the minimum bounding rectangle. In a second step the nDEM area of this rectangle is rotated and cut to a single building nDEM B_i . In a second step for simple modeling a “no-building-mask” is calculated for each B_i by fetching the mean height m_i of B_i and creating a binary mask for $B_i < m_i$. For all segments from this binary “no-building-mask” are in turn the minimum bounding rectangles are calculated and these are subtracted from the buildings bounding rectangle. Ideally this procedure is iteratively repeated – always inverting the mask and cutting/adding the new rectangles from the previous.¹⁶ The 1304 automatically extracted buildings are shown in fig. 10 (center).

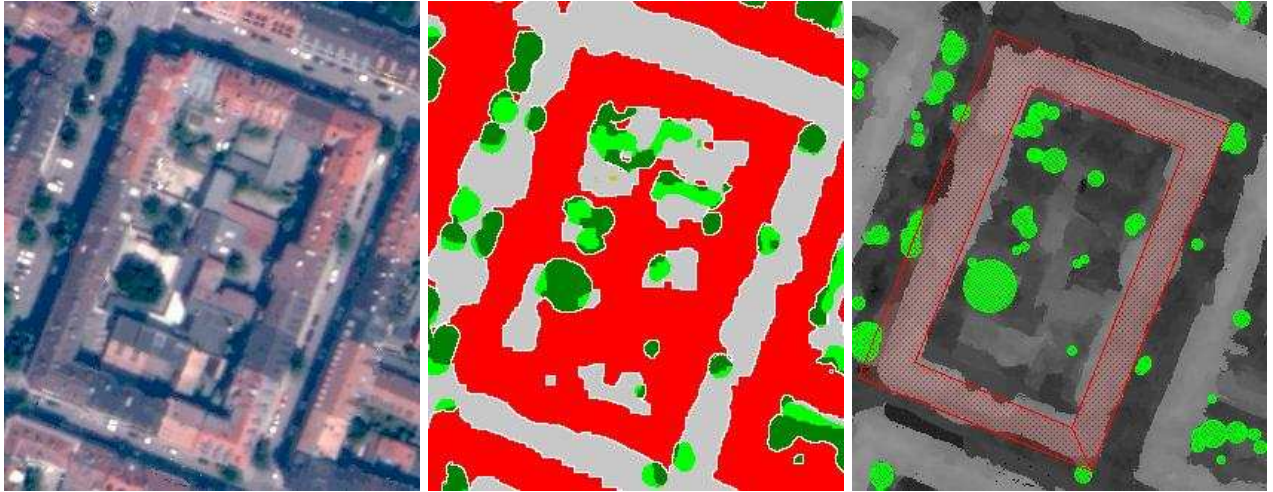


Figure 9. Small section 280×330 px of the test area showing the panch sharpened ortho image (left), the spectral classification (center) and the automatically derived tree- (green) and building-objects (red) overlaid on the DSM (right)



Figure 10. 41118 trees (left) and 1304 buildings (center) overlaid on the DSM as LOD1 vector objects in the $3000 \text{ m} \times 3000 \text{ m}$ test scene, right: trees and buildings combined

3. CONCLUSION AND OUTLOOK

In this paper we presented an automatic processing chain for extracting urban objects from single VHR satellite stereo pairs without need of any additional information. First the so called height map is derived using the dense stereo matching method “semi global matching” on the panchromatic stereo image pair. A newly method for filling the intermediate height maps based on the original panchromatic and multispectral input imagery is presented followed by the orthorectification of the height map to a digital surface model (DSM) and the panch sharpened imagery.

In the second section the digital terrain model and the nDEM is derived from the DSM and a fuzzy-based multispectral classification is applied to the panch sharpened ortho image. These classifications are used for the object extraction and modeling in the third part of the paper. Fig. 11 shows a resulting 3D view of a small section from the test data including buildings, trees and a smoothly rolling DTM.

Further investigations will be necessary on a better building reconstruction up to LOD2 by segmentation of the buildings based on the nDEM and the ortho imagery as also the implementation of the extraction and modeling of more urban objects like water bodies or bridges.

REFERENCES

- [1] “DigitalGlobe.” <http://www.digitalglobe.com/> (3 2014). (accessed 01/2014).



Figure 11. Small section of the test area including LOD1 buildings, trees and the DTM

- [2] Hirschmüller, H., "Stereo Processing by Semiglobal Matching and Mutual Information," *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE* **30/2**, 328–341 (2008).
- [3] d'Angelo, P., Schwind, P., Krauß, T., Barner, F., and Reinartz, P., "Automated DSM based Georeferencing of CARTOSAT-1 Stereo Scenes," in [*International Archives of Photogrammetry and Remote Sensing*], *International Archives of Photogrammetry and Remote Sensing* **39** (4 2009).
- [4] Jacobsen, K., Büyüksalih, G., and Topan, H., "Geometric Models for the orientation of high resolution optical satellite sensors," in [*ISPRS Workshop, Hannover*], *ISPRS* **36 (1/W3)** (2005).
- [5] Grodecki, J., Dial, G., and Lutes, J., "Mathematical Model for 3D feature extraction from multiple satellite images described by RPCs," in [*ASPRS Annual Conference Proceedings, Denver, Colorado*], *ASPRS Annual Conference Proceedings, Denver, Colorado* (2004).
- [6] Morgan, M., *Epipolar Resampling of Linear Array Scanner Scenes*, PhD thesis, University of Calgary, Canada (5 2004). UCGE Reports Nr. 20193.
- [7] Du, Q., Younan, N., King, R., and Shah, V., "On the performance evaluation of pan-sharpening techniques," *Geoscience and Remote Sensing Letters, IEEE* **4**, 518 – 522 (10 2007).
- [8] Krauß, T., Sirmacek, B., Arefi, H., and Reinartz, P., "Fusing stereo and multispectral data from WorldView-2 for urban modeling," in [*Proc. SPIE 8390*], *Proc. SPIE 8390* **83901X**, 2012 (4 2012).
- [9] Krauß, T., "Preprocessing of Satellite Data for Urban Object Extraction," in [*ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*], *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* , 115–120 (3 2015).
- [10] Weidner, U. and Förstner, W., "Towards automatic building extraction from high resolution digital elevation models," *ISPRS* **50 (4)**, 38–49 (1995).
- [11] Arefi, H., Engels, J., Hahn, M., and Mayer, H., "Automatic DTM generation from laser-scanning data in residential hilly area," *Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* **36(4)** (2007).

- [12] Krauß, T. and Reinartz, P., “Urban object detection using a fusion approach of dense urban digital surface models and VHR optical satellite stereo data,” in [*ISPRS Istanbul Workshop 2010, WG I/4 Modeling of optical airborne and space borne sensors*], *ISPRS* **39** (10 2010).
- [13] Krauß, T., Lehner, M., and Reinartz, P., “Generation of coarse 3D models of urban areas from high resolution stereo satellite images,” in [*IAPRS*], *IAPRS* **37** (7 2008).
- [14] “CityGML-Definition.” <http://www.opengeospatial.org/standards/citygml> (11 2008). (accessed 01/2014).
- [15] Sirmacek, B., dAngelo, P., and Reinartz, P., “Detecting complex building shapes in panchromatic satellite images for digital elevation model enhancement,” in [*ISPRS Istanbul Workshop 2010 on Modeling of optical airborne and spaceborne Sensors*], *IAPRS* **38 1/W4** (10 2010).
- [16] Arefi, H. and Reinartz, P., “Building reconstruction from Worldview DEM using image information,” in [*SMPR2011*], *SMPR2011* (5 2011).